

**text extraction**

**Main plan:**

extract text and tables from a text file or html file. For the tables, we want to have a xlsx file including all the tables written in different sheets, the rows and columns of tables be well detected. For the text, we want to create a xlsx file including the page information, the type of the extracted text and the found text in the report.

**Accomplished work:**

We write the code to analyze the structure of the html code, extract the information according to html code and convert the data to the form we need.

Finally, we get the file we wanted.